Bayesian Analysis of Loss Ratios Using the Reversible Jump Algorithm

G. O. Brown†

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge CB3 0WB, UK

S. P. Brooks

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge CB3 0WB, UK

Summary. In this paper we consider the problem of model choice for a set of insurance loss ratios. We use a reversible jump algorithm for our model discrimination and show how the vanilla reversible jump algorithm can be improved on using recent methodological advances in reversible jump computation.

†Corresponding author, Email:gob20@statslab.cam.ac.uk

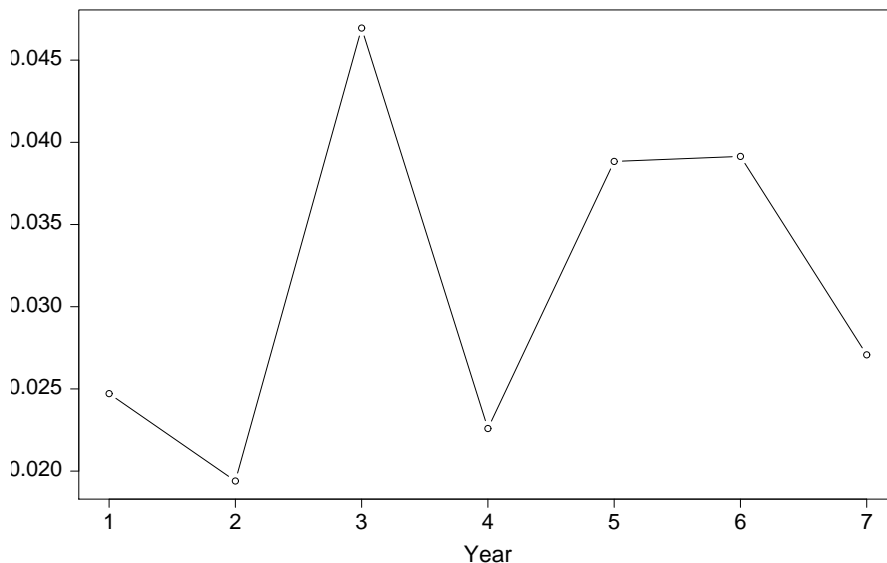**Fig. 1.** A plot of the loss ratios against year.

1. Introduction

In traditional insurance settings model selection and uncertainty are usually not treated, even though model selection problems have been actively researched in statistics. Recently this shortcoming has been addressed by several authors: Cairns (2000) in general insurance and risk theory; Keatinge (1999) in estimating the number of components in a mixture of exponentials for estimating claims amounts; also Harris (1999) considers the problem of model selection for vector autoregression for financial time series. In the field of credibility theory, Bühlmann and Bühlmann (1999) considers the selection of variables in certain regression credibility models.

In this paper we consider the problem of parameter estimation and model selection in the analysis of workers' compensation loss ratios. This paper is motivated by the analysis of data consisting of workers' compensation loss ratios arising over a seven year period. The data are part of a set containing frequency counts on workers' compensation insurance. The number of claims against the workers' compensation insurance scheme is recorded, together with the corresponding exposure values. The exposures are scaled payroll totals and provide a measure of the size of the exposed group.

The model we fit to the data is described in Section 2, we then introduce two additional models, both of which are sub-models of the first. Using the reversible jump method described in Section 4 we discriminate between the three models. We also use the efficient proposals method of Brooks et al. (2003) to derive proposals for our reversible jump updates. The results are compared with the pilot-tuned vanilla reversible jump of Green (1995).

2. The Data and Model

We denote the number of claims for year j by L_j and the corresponding exposure values by E_j for $j = 1, \ldots, n$. For this particular dataset we have $n = 7$. The loss ratios which we propose to model are then defined as the number of losses per unit exposure and will be denoted by R_j, where $R_j = L_j / E_j$. Let R^n denote the collective loss ratios and E^n denote the collective exposure values. Here, we use a hierarchical normal model to describe the loss ratios, so that

$$R_j \sim \mathcal{N}\left(\alpha_j, (\sigma E_j)^{-1}\right) \ j = 1, \ldots, n \tag{1}$$

where α_j denotes some underlying time-varying process which describes the progression of ratio level over time. Here, we follow Klugman (1992) and adopt the following model for the α_j process

$$\alpha_j \sim \mathcal{N}\left(\rho \alpha_{j-1} + (1-\rho)\eta, \tau^{-1}\right) \ j = 1, \ldots, n. \tag{2}$$

For α_0, ρ, and η we use standard normal $\mathcal{N}(0,1)$ priors and for the precision (inverse variance) parameters σ and τ we use Gamma(a,b) priors. The literature provides empirical evidence to support the introduction of this model for describing loss ratios (Ledolter et al., 1991) and a simple plot of the data in Figure 1 confirms that the observed behaviour can be described by a model of this sort. However, one disadvantage of this model is that whereas the loss ratios are always non-negative, the normal model has support extending across the entire real line so that negative values could, in theory, occur. One way around this would be to restrict the normal model in (1) to loss ratios within the region $[0, \infty)$ and/or to impose similar restrictions on the α_j process. These restrictions are very easily implemented as a trivial extension of the scheme we describe here but, since by adopting the more general model, serious failures in the ability of the model to describe the observed data can be detected when negative estimates are obtained, the more general model provides a useful check for the adequacy of our modelling scheme.

3. A Gibbs Sampling Algorithm

For the model presented in Equations (1) and (2) we use Gibbs updates to obtain samples from the posterior distribution of the model parameters. The full joint posterior distribution of all the model parameters is given by

$$\pi(\alpha, \tau, \sigma, \rho, \eta | R^n, E^n) \propto L(R^n | \alpha, \sigma, E^n) p(\alpha | \rho, \eta, \tau, \alpha_0) p(\sigma) p(\tau) p(\rho) p(\eta) p(\alpha_0),$$

where α denotes the collection $(\alpha_1, \cdots, \alpha_n)$, $p(\cdot)$ denotes the prior distribution for the corresponding parameter and the likelihood term

$$L(R^n | \alpha, \sigma, E^n) = \prod_{j=1}^{n} f(R_j | \alpha, \sigma, E^n).$$

We now derive the full posterior conditional for each of the model parameters in turn. These conditional distributions will then be used to implement a Gibbs update algorithm. The full posterior conditional distribution for σ is

$$\begin{aligned}
\pi(\sigma | \alpha) &\propto p(\sigma) L(R^n | \alpha, \sigma, E^n) \\
&\propto \sigma^{a + \frac{n}{2} - 1} \exp\left\{-\sigma\left(b + \frac{1}{2}\sum_{j=1}^{n} E_j (R_j - \alpha_j)^2\right)\right\},
\end{aligned}$$

which is a gamma distribution with shape parameter $a+\frac{n}{2}$ and scale parameter $b+\frac{1}{2}\sum_{j=1}^{n}E_j(R_j-\alpha_j)^2$. The full posterior conditional for τ is

$$\pi(\tau|\alpha,\rho,\eta) \propto p(\tau)p(\alpha|\alpha_0,\rho,\eta,\tau)$$
$$\propto \tau^{a+\frac{n}{2}-1}\exp\left\{-\tau\left(b+\frac{1}{2}\sum_{j=1}^{n}(\alpha_j-\rho\alpha_{j-1}-(1-\rho)\eta)^2\right)\right\},$$

which is a gamma distribution with shape parameter $a+\frac{n}{2}$ and scale parameter $b+\frac{1}{2}\sum_{j=1}^{n}(\alpha_j-\rho\alpha_{j-1}-(1-\rho)\eta)^2$. The full posterior conditional distribution for ρ is

$$\pi(\rho|\alpha,\eta,\tau) \propto p(\rho)p(\alpha|\alpha_0,\rho,\eta)$$
$$\propto \exp\left\{-\frac{(1+\tau\sum_{j=1}^{n}(\eta-\alpha_{j-1})^2)}{2}\left(\rho-\frac{\tau\sum_{j=1}^{n}(\eta-\alpha_j)(\eta-\alpha_{j-1})}{1+\tau\sum_{j=1}^{n}(\eta-\alpha_{j-1})^2}\right)^2\right\},$$

which is a normal distribution with mean

$$(1+\tau\sum_{j=1}^{n}(\eta-\alpha_{j-1})^2)^{-1}(\tau\sum_{j=1}^{n}(\eta-\alpha_j)(\eta-\alpha_{j-1})),$$

and variance $(1+\tau\sum_{j=1}^{n}(\eta-\alpha_{j-1})^2)^{-1}$. The full posterior conditional distribution for η is

$$\pi(\eta|\alpha,\rho,\tau) \propto p(\eta)p(\alpha|\alpha_0,\rho,\eta)$$
$$\propto \exp\left\{-\frac{1+n\tau(1-\rho)^2}{2}\left(\eta-\frac{\tau(1-\rho)\sum_{j=1}^{n}(\alpha_j-\rho\alpha_{j-1})}{1+n\tau(1-\rho)^2}\right)^2\right\},$$

which is a normal distribution with mean

$$(1+n\tau(1-\rho)^2)^{-1}(\tau(1-\rho)\sum_{j=1}^{n}(\alpha_j-\rho\alpha_{j-1})),$$

and variance $(1+n\tau(1-\rho)^2)^{-1}$. The full posterior conditional distribution for α_j is

$$\pi(\alpha_j|\alpha_{(j)},\rho,\eta,\sigma,\tau) \propto$$
$$\begin{cases} p(\alpha_j)p(\alpha_{j+1}|\alpha_j,\rho,\eta,\tau) & j=0 \\ p(\alpha_j|\alpha_{j-1},\rho,\eta,\tau)p(\alpha_{j+1}|\alpha_j,\rho,\eta,\tau)p(R_j|\alpha_j,\sigma) & j=1,\ldots,n-1 \\ p(\alpha_j|\alpha_{j-1},\rho,\eta,\tau)p(R_j|\alpha_j,\sigma) & j=n \end{cases},$$

which is a normal distribution with mean m_j and variance v_j where

$$m_j = \begin{cases} \dfrac{(\rho\tau(\alpha_{j+1}-(1-\rho)\eta))}{(1+\rho^2\tau)} & j=0 \\ \dfrac{(\rho\tau(\alpha_{j+1}-(1-\rho)\eta)+\tau(\rho\alpha_{j-1}+(1-\rho)\eta)+\sigma E_j R_j)}{(\rho^2\tau+\tau+\sigma E_j)} & j=1,\ldots,n-1 \\ \dfrac{(\tau(\rho\alpha_{j-1}+(1-\rho)\eta)+\sigma E_j R_j)}{(\tau+\sigma E_j)} & j=n, \end{cases}$$

and

$$
v_j = \begin{cases} \dfrac{1}{(1+\rho^2\tau)} & j = 0 \\[2ex] \dfrac{1}{(\rho^2\tau+\tau+\sigma E_j)} & j = 1,\ldots,n-1, \\[2ex] \dfrac{1}{(\tau+\sigma E_j)} & j = n. \end{cases}
$$

These conditional distributions are then used to simulate a dependent sample from the posterior distribution of the model parameters given the data R^n by sampling each in turn within each iteration. All the posterior conditionals are standard distributions, hence there are no difficulties in simulating from them. We could construct a more general Metropolis type algorithm which updates the parameters α, ρ and η at the same time. This would necessitate introducing an acceptance/rejection stage to ensure stationarity.

3.1.   Simulation Results

The Gibbs model above was implemented with $a = 0.001$ and $b = 0.001$ so that the precision parameters σ and τ have vague, flat priors. The posterior means and 95% highest posterior density (HPD) intervals are shown in Table 1 and trace plots of the model parameters are shown in Figure 4. The 95% HPD interval is the smallest region of the parameter space which contains 95% of the posterior probability mass of the parameter. A plot of the marginal posterior of ρ reveals that its density is bimodal with one mode near 0 and another at $\rho = 1$. The posterior density of rho is shown in Figure 2. Even though the 95% HPD interval consists of only one interval, a corresponding 90% HPD interval is actually a union of two disjoint intervals, each containing one of the two modes. The effect of the bimodality of ρ can be seen by the wide intervals for the parameters α_0 and η.

A possible explanation is that the posterior conditional of α_0 is the same as its prior density, since there is no need for α_0 if ρ is identically 0. Similarly when ρ is close to the mode at 1, the conditional posterior of η is almost identical to its prior density. Consequently, these two parameters are being sampled from two distinct posterior densities corresponding to whether ρ is close to 0 or 1. The reason for the bimodality of ρ is not entirely clear, the model may be overparameterised since we are fitting 12 parameters to 7 data points. To observe the effect of the number of parameters we can reduce the effective number of parameters being fitted by integrating out the nuisance parameters σ and τ, then re-fitting the model and observing any differences. The results of this new implementations are identical to the first implementation with both σ and τ included, as we show in the next section.

3.2.   Integrating out the Variance Parameters

Papaspiliopoulos et al. (2003) shows that for Gaussian models similar to that described in Equations 1 and (2) the convergence properties are largely determined by the values of the variance components. In this Section we redo the analysis, however this time we integrate out the variance parameters σ and τ. This results in fewer parameters to be estimated, but as the results show, it also increases the autocorrelation of the other parameters. Also the complexity of the model has been reduced. The form of the conditional posteriors, however, has been made more complex. We use a random walk Metropolis algorithm to simulate from the posterior distribution of the remaining unknown parameters $\alpha_0, \alpha_1, \ldots, \alpha_n, \rho, \eta$. The results show that the posterior estimate of ρ is still bimodal.

6

Table 1. Posterior means and 95% HPD Intervals for the model parameters with $a = b = c = 0.001$.

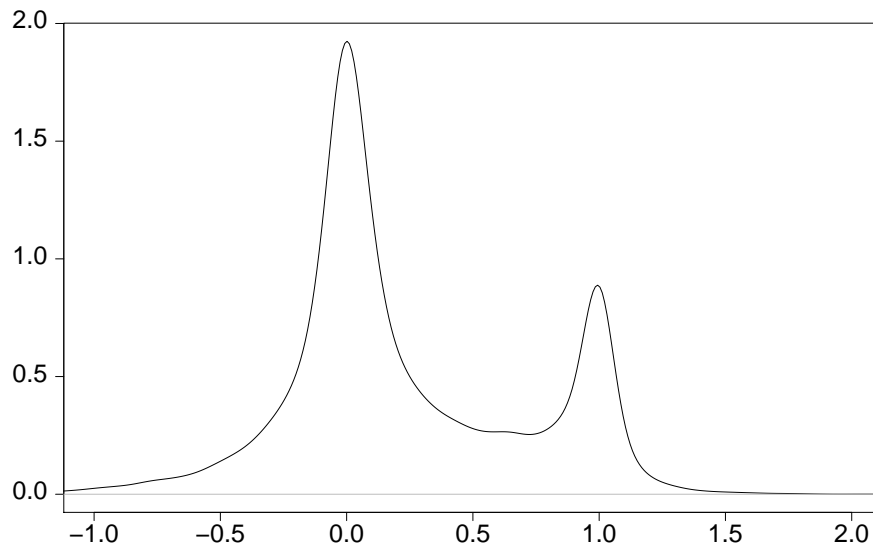| Parameter	estimate	95% HPD Interval
α_0	0.0167	(-1.1188, 1.0619)
α_1	0.0256	(-0.0311, 0.0817)
α_2	0.0246	(-0.0211, 0.0708)
α_3	0.0398	(-0.0067, 0.0846)
α_4	0.0271	(-0.0165, 0.0712)
α_5	0.0362	(-0.0064, 0.0792)
α_6	0.0364	(-0.0072, 0.0785)
α_7	0.0296	(-0.0170, 0.0772)
ρ	0.220	(-0.483, 1.154)
η	0.0315	(-0.3831, 0.4477)
σ	1014.9	(0.0, 2634.6)
τ	1371.2	(0.0, 3330.0)

Fig. 2. Posterior density of ρ.

If $\sigma \sim \text{Gamma}(a_1, b_1)$ and $\tau \sim \text{Gamma}(a_2, b_2)$, recall that in Section 3 we showed that the posterior conditionals of σ and τ are of the form of Gamma densities, and both are independent of each other, since the posterior conditional of σ does not depend on τ and likewise that of τ does not depend on σ.

Since also the posterior conditionals are of standard form, we can integrate out these two parameters leaving a density involving only the other parameters. Using the fact that the posterior conditionals for σ and τ are standard Gamma densities, we can show that

$$\pi(\alpha, \alpha_0, \rho, \eta | R^n) = \int \pi(\alpha, \alpha_0, \rho, \eta, \sigma, \tau | R^n) \, d\sigma d\tau,$$

so that

$$\pi(\alpha, \alpha_0, \rho, \eta | R^n) \propto p(\rho)p(\eta)p(\alpha_0) \times \left(b_1 + \tfrac{1}{2}\sum E_j(\alpha_j - R_j)^2\right)^{-(a_1 + n/2)} \times$$
$$\left(b_2 + \tfrac{1}{2}\sum(\alpha_j - \rho\alpha_{j-1} - (1-\rho)\eta)^2\right)^{-(a_2 + n/2)} \quad (3)$$

where $\pi(\alpha, \alpha_0, \rho, \eta | R^n)$ is the posterior density of α, α_0, ρ and η given the data.

Now given (3) the following posterior conditionals are readily observed

$$\pi(\rho | \alpha, \alpha_0, \eta) \propto p(\rho) \left(b_2 + \tfrac{1}{2}\sum(\alpha_j - \rho\alpha_{j-1} - (1-\rho)\eta)^2\right)^{-(a_2 + n/2)},$$

and

$$\pi(\eta | \alpha, \alpha_0, \rho) \propto p(\eta) \left(b_2 + \tfrac{1}{2}\sum(\alpha_j - \rho\alpha_{j-1} - (1-\rho)\eta)^2\right)^{-(a_2 + n/2)},$$

also

$$\pi(\alpha, \alpha_0 | \rho, \eta) \propto p(\alpha_0) \left(b_1 + \tfrac{1}{2}\sum E_j(\alpha_j - R_j)^2\right)^{-(a_1 + n/2)} \times$$
$$\left(b_2 + \tfrac{1}{2}\sum(\alpha_j - \rho\alpha_{j-1} - (1-\rho)\eta)^2\right)^{-(a_2 + n/2)}.$$

We use this scheme because our attempts to update ρ, η and α as one block using a 10-variate normal distribution centred at the current values did not work very well.

These are all non-standard densities and to implement this model we used a Gibbs updating scheme with random walk Metropolis algorithms for η, ρ and α_0 with uniform distributions centred at the current values. For ρ, η and α_0, the width of the proposal interval was determined by fine tuning an initial run until the acceptance rates were 0.27, 0.15 and 0.29, respectively. For $\alpha_1, \ldots, \alpha_7$ we used a 7-variable normal density as the proposal for a random walk Metropolis algorithm centred at the current values of these parameters. The covariance matrix for this proposal distribution was determined from an initial run from which we computed the covariance of the parameters $\alpha_1, \ldots, \alpha_7$. With this covariance matrix the acceptance rate of the Metropolis algorithm is 0.15, this is smaller than would be ideal (Roberts and Rosenthal, 1998, 2001).

The results for this model are shown in Table 2, they are similar to those in Table 1. The main difference here is that the 95% HPD intervals for α_0 and η are now smaller and more concentrated around the posterior means, which represents an improvement on those given in Table 1.

An important diagnostic tool in MCMC modelling is the autocorrelation plot of the parameter of interest. Figures 3 and 5 show the autocorrelation functions for the parameters of interest in our model. For the full

Table 2. Posterior means and 95% HPD Intervals for the model parameters, after integrating out the variance parameters with $a_1 = a_2 = b_1 = b_2 = 0.001$.

Parameter	Estimate	95% HPD Interval
α_0	0.01597	(-1.1332, 1.0953)
α_1	0.02551	(-0.0312, 0.0830)
α_2	0.02439	(-0.0204, 0.0706)
α_3	0.03987	(-0.0063, 0.0842)
α_4	0.02703	(-0.0161, 0.0705)
α_5	0.03630	(-0.0086, 0.0781)
α_6	0.03654	(-0.0072, 0.0776)
α_7	0.02942	(-0.0169, 0.0777)
ρ	0.21410	(-0.5585, 1.1177)
η	0.02775	(-0.4214, 0.4264)

implementation the autocorrelation values are essentially zero at lags greater than 5, except for ρ where lags up to 25 are large. The implementation with the inverse-variance parameters integrated out does not appear to be better than the full implementation. The autocorrelations for this implementation are bigger than those of the full implementation at all lags and are significant up to lag 10, excepting for ρ which has autocorrelation significant up to lag 15. The autocorrelation plots and trace plots for the implementation with the variance parameters integrated out are shown in Figures 6 and 5, respectively. This could be indicating poor mixing of the Metropolis algorithm due to the complexity of the terms in Equation (3) and small Metropolis acceptance rates. The small Metropolis rates could also be indicating that the proposal variances need to be smaller so that proposed values are closer to the current values and will have a greater chance of being accepted.

These results suggest that fitting models with $\rho = 0$ or $\rho = 1$ should give better description of the data. i.e. $\rho = 0$ and $\rho = 1$ might be plausible alternative models. In the remainder of this paper we examine in more detail the simplified models with $\rho = 0$ and $\rho = 1$, and how to discriminate between them. The model with $\rho = 0$ means that a simple variance components model will be enough to describe the data whereas the model with $\rho = 1$ means a simple autoregressive model with Normal errors will be adequate to describe the data.

4. The Reversible Jump Algorithm

In this section we discuss trans-dimensional algorithms. The algorithms we have discussed before, notably the Metropolis–Hastings algorithm and the Gibbs sampling algorithm, cannot be used to simulate Markov chains where the dimension of the state vector can change at each iteration. This situation arises particularly in model selection problems where there are competing models, and where the size of the parameter vector is allowed to vary between models.

In this context the distribution of interest is defined jointly over both parameter and model space. Several authors have proposed simulation methods to construct Markov chains which can explore such state spaces. These include the product space formulation given in Carlin and Chib (1995), the reversible jump (RJMCMC) algorithm of Green (1995), the jump diffusion method of Grenander and Miller (1994), and Phillips and Smith (1996) and the continuous time birth-death method of Stephens (2000). Also for particular problems involving the size of the regression vector in regression analysis there is the stochastic search variable selection method of George and McCulloch (1993). In the remainder of this section we describe the
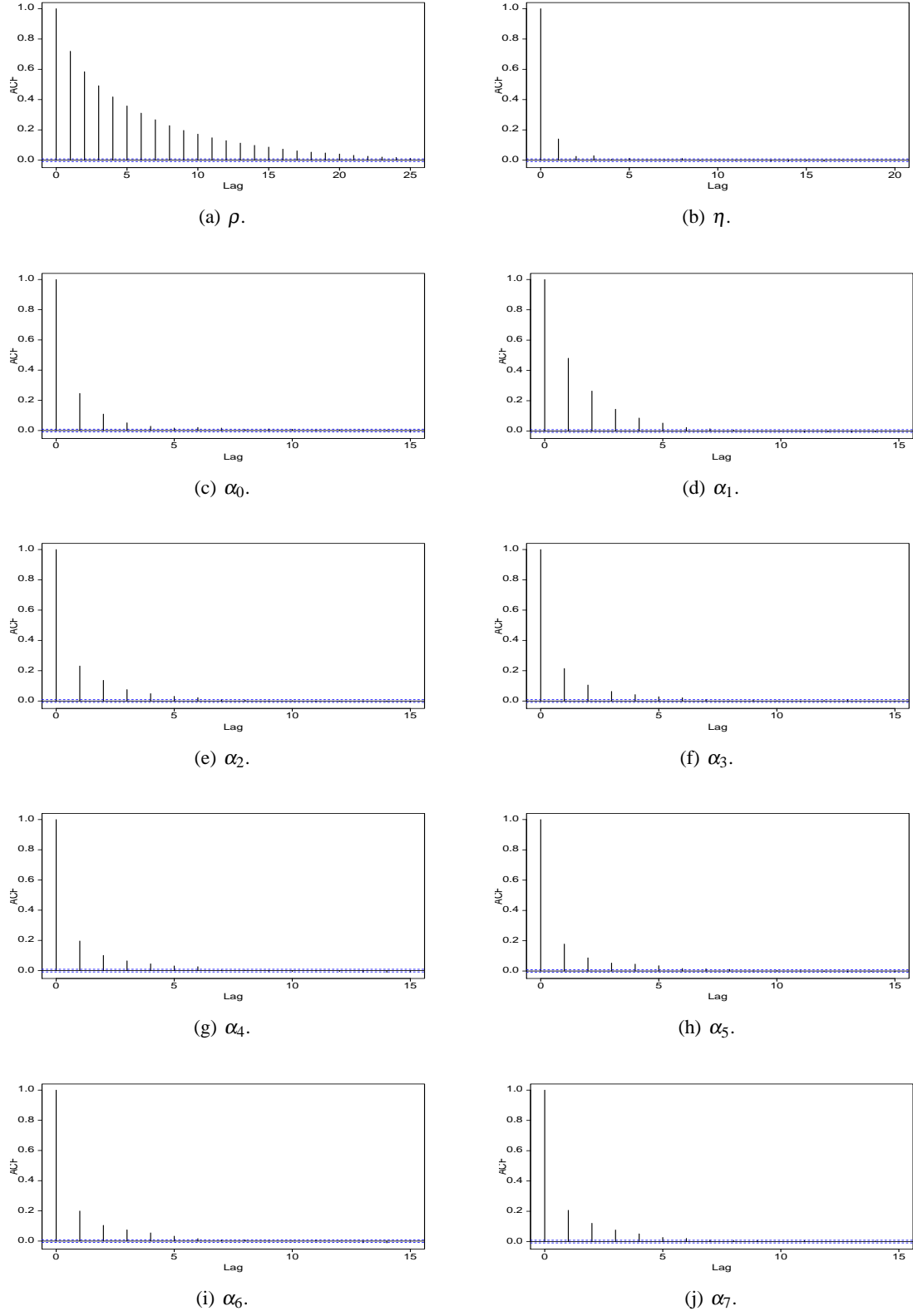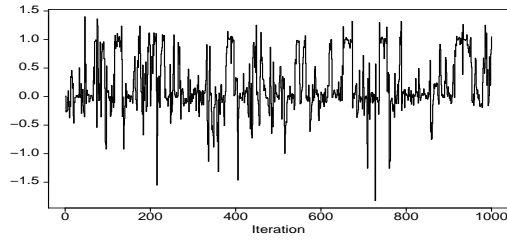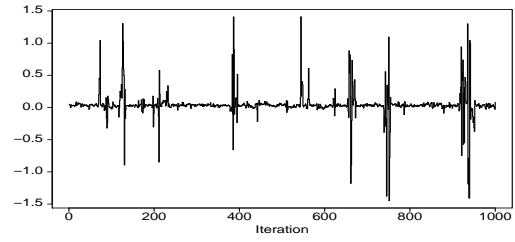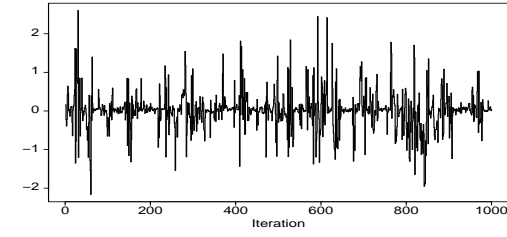
(a) ρ.

(b) η.

(c) α_0.

(d) α_1.

(e) α_2.

(f) α_3.

(g) α_4.

(h) α_5.

(i) α_6.

(j) α_7.

Fig. 3. Autocorrelation plots for ρ, η and α.

(a) Trace plot of ρ.

(b) Trace plot of η.

(c) Trace plot of α_0.
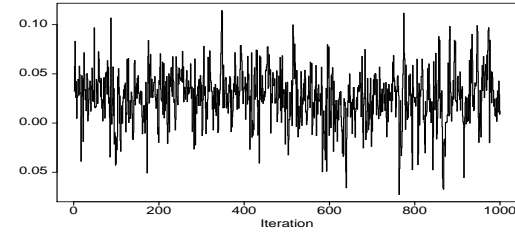
(d) Trace plot of α_1.
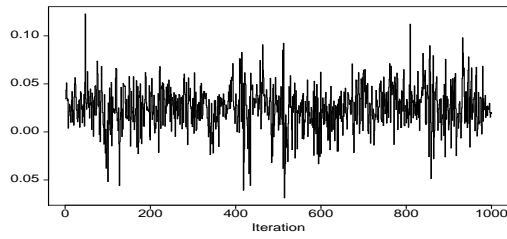
(e) Trace plot of α_2.
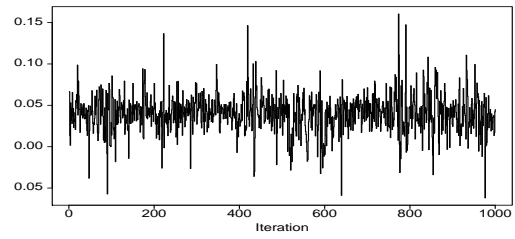
(f) Trace plot of α_3.
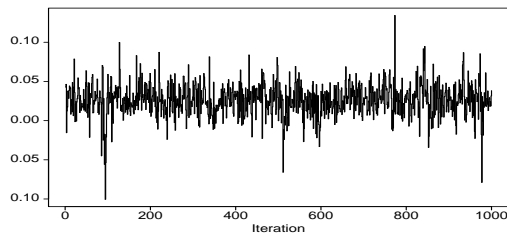
(g) Trace plot of α_4.
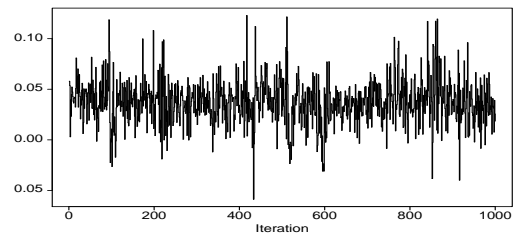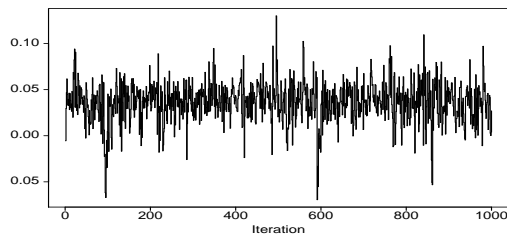
(h) Trace plot of α_5.
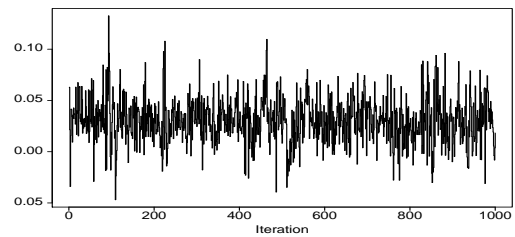
(i) Trace plot of α_6.

(j) Trace plot of α_7.

Fig. 4. Trace plots of model parameters.
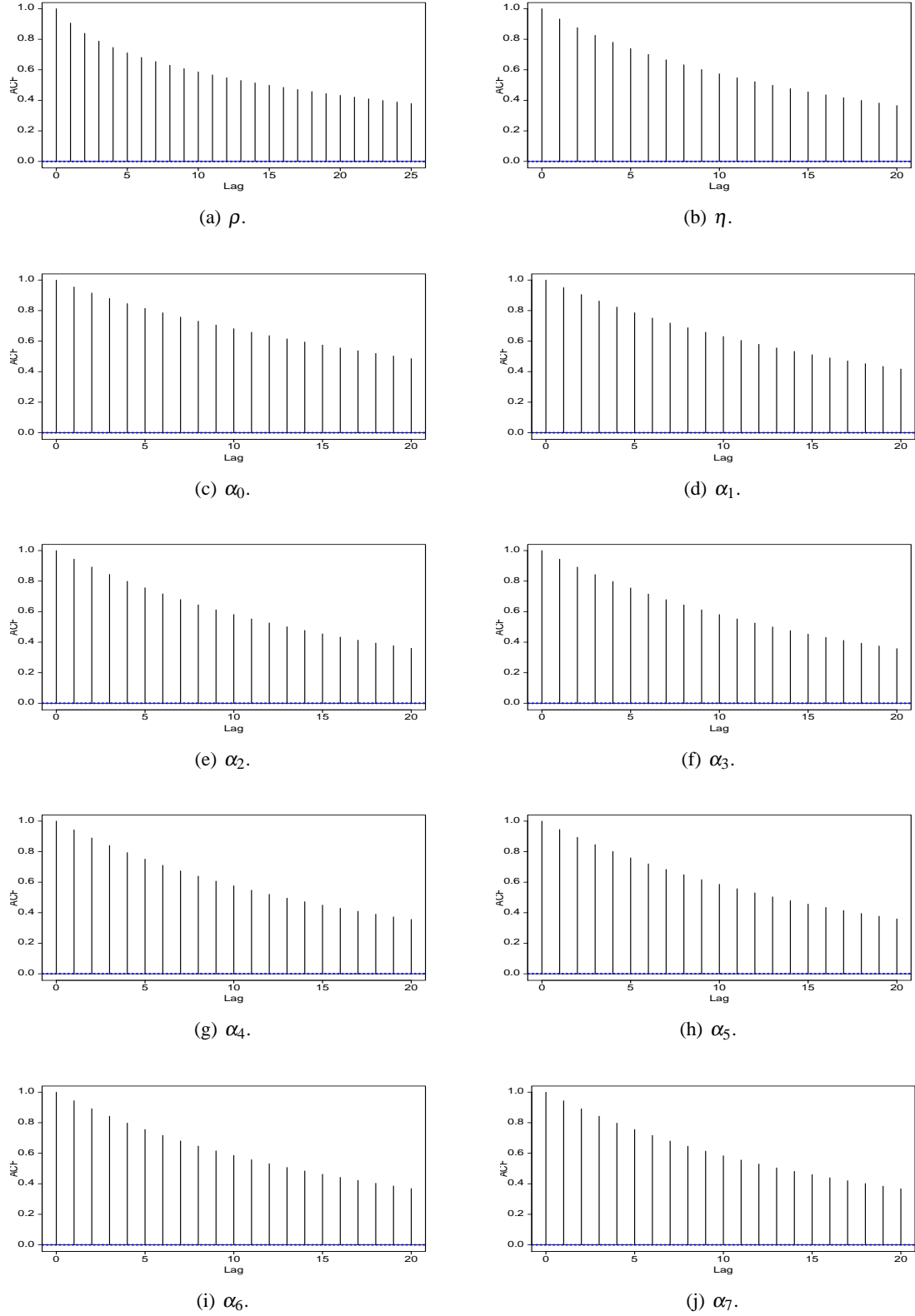
Fig. 5. Autocorrelation plots for ρ, η and α with inverse-variance parameters σ and τ integrated out.
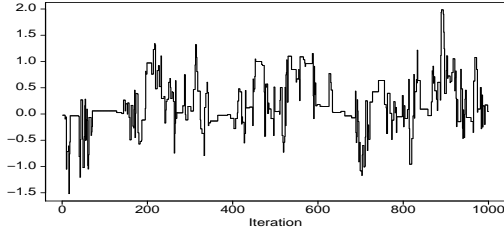
(a) Trace plot of ρ.
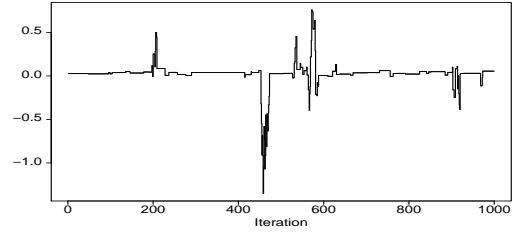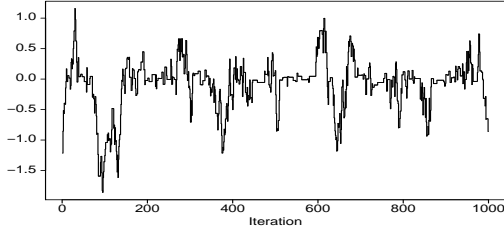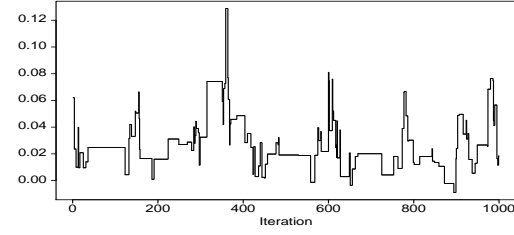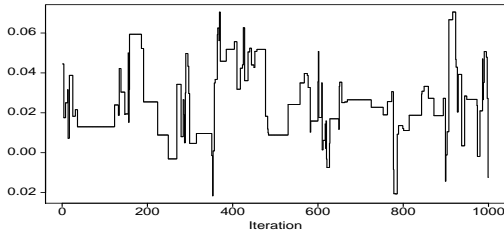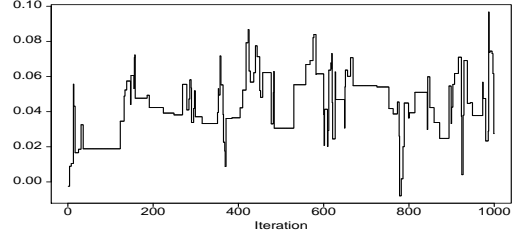
(b) Trace plot of η.

(c) Trace plot of α_0.

(d) Trace plot of α_1.

(e) Trace plot of α_2.

(f) Trace plot of α_3.
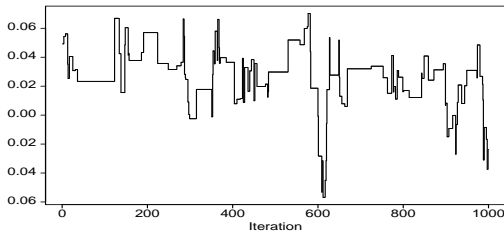
(g) Trace plot of α_4.
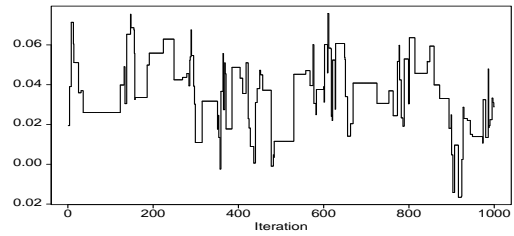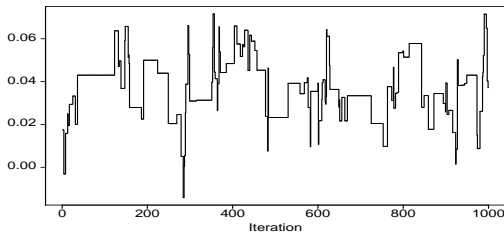
(h) Trace plot of α_5.
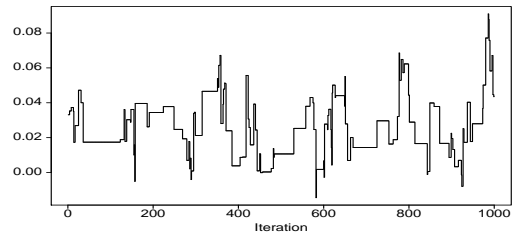
(i) Trace plot of α_6.

(j) Trace plot of α_7.

Fig. 6. Trace plots of model parameters, after integrating out the inverse-variance parameters.

reversible jump method of Green (1995). In practice trans–dimensional algorithms work by updating model parameters for the current model then proposing to change models with some specified probability.

The Reversible jump algorithm represents an extension of the Metropolis–Hastings algorithm. We assume there is a countable collection of candidate models, indexed by $M \in \mathscr{M} = \{M_1, M_2, \dots, M_k\}$. We further assume that for each model M_i, there exists an unknown parameter vector $\theta_i \in \mathbb{R}^{n_i}$ where n_i, the dimension of the parameter vector, can vary with i.

Typically we are interested in finding which models have the greatest posterior probabilities and also estimates of the parameters. Thus the unknowns in this modelling scenario will include the model index M_i as well as the parameter vector θ_i. We assume that the models and corresponding parameter vectors have a joint density $\pi(M_i, \theta_i)$. The reversible jump algorithm constructs a reversible Markov chain on the state space $\mathscr{M} \times \bigcup_{M_i \in \mathscr{M}} \mathbb{R}^{n_i}$ which has π as its stationary distribution (Green, 1995). In many instances, and in particular for Bayesian problems this joint distribution is of the form

$$\pi(M_i, \theta_i) = \pi(M_i, \theta_i | X) \propto L(X | M_i, \theta_i) \, p(M_i, \theta_i),$$

where the prior on (M_i, θ_i) is often of the form

$$p(M_i, \theta_i) = p(\theta_i | M_i) \, p(M_i)$$

with $p(M_i)$ being the density of some counting distribution.

Suppose now that we are at model M_i and a move to model M_j is proposed with probability r_{ij}. The corresponding move from θ_i to θ_j is achieved by using a deterministic transformation h_{ij}, such that

$$(\theta_j, v) = h_{ij}(\theta_i, u), \qquad (4)$$

where u and v are random variables introduced to ensure dimension matching necessary for reversibility. To ensure dimension matching we must have

$$\dim(\theta_j) + \dim(v) = \dim(\theta_i) + \dim(u).$$

For discussions about possible choices for the function h_{ij} we refer the reader to Green (1995), and Brooks et al. (2003). Let

$$A(\theta_i, \theta_j) = \frac{\pi(M_j, \theta_j)}{\pi(M_i, \theta_i)} \frac{q(v)}{q(u)} \frac{r_{ji}}{r_{ij}} \left| \frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)} \right| \qquad (5)$$

then the acceptance probability for a proposed move from model (M_i, θ_i) to model (M_j, θ_j) is

$$\min \left\{ 1, A(\theta_i, \theta_j) \right\}$$

where $q(u)$ and $q(v)$ are the respective proposal densities for u and v, and $|\partial h_{ij}(\theta_i, u)/\partial(\theta_i, u)|$ is the Jacobian of the transformation induced by h_{ij}. Green (1995) shows that the algorithm with acceptance probability given above simulates a Markov chain which is reversible and follows from the detailed balance equation

$$\pi(M_i, \theta_i)q(u)r_{ij} = \pi(M_j, \theta_j)q(v)r_{ji} \left| \frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)} \right|.$$

Detailed balance is necessary to ensure reversibility and is a sufficient condition for the existence of a unique stationary distribution. For the reverse move from model M_j to model M_i it is easy to see that the transformation used is $(\theta_i, u) = h_{ij}^{-1}(\theta_j, v)$ and the acceptance probability for such a move is

$$\min\left\{1, \frac{\pi(M_i, \theta_i)}{\pi(M_j, \theta_j)} \frac{q(u)}{q(v)} \frac{r_{ij}}{r_{ji}} \left|\frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)}\right|^{-1}\right\} = \min\left\{1, A(\theta_i, \theta_j)^{-1}\right\}.$$

For inference regarding which model has the greater posterior probability we can base our analysis on a realisation of the Markov chain constructed above. The marginal posterior probability of model M_i

$$\pi(M_i|X) = \frac{p(M_i)f(X|M_i)}{\sum_{M_j \in \mathcal{M}} p(M_j)f(X|M_j)},$$

where

$$f(X|M_i) = \int L(X|M_i, \theta_i)p(\theta_i|M_i)\,d\theta_i$$

is the marginal density of the data after integrating over the unknown parameters θ. In practice we estimate $\pi(M_i|X)$ by counting the number of times the Markov chain visits model M_i in a single long run after reaching stationarity. These between model moves described in this section are also augmented with within model Gibbs updates as given in Section 3 to update model parameters.

4.1. Efficient Proposals

In practice the between model moves can be small resulting in poor mixing of the resulting Markov chain. In this section we discuss recent attempts at improving between model moves by increasing the acceptance probabilities for such moves. Several authors have addressed this problem including Troughton and Godsill (1997), Giudici and Roberts (1998), Godsill (2001), Rotondi (2002), and Al-Awadhi et al. (2004). Green and Mira (2001) proposes an algorithm so that when between model moves are first rejected, a second attempt is made. This algorithm allows for a different proposal to generated from a new distribution, that is allowed to depend on the previously rejected proposal. Methods to improve mixing of reversible jump chains have also been proposed by Green (2002) and Brooks et al. (2003), which has been extended by Ehlers and Brooks (2002).

A general strategy proposed by Brooks et al. (2003) and extended to more general cases by Ehlers and Brooks (2002) is based on making the term $A_{ij}(\theta_i, \theta_j)$ in the acceptance probability for between model moves given in Equation (5) as close as possible to 1. The motivating reason for this is that if we make this term as close as possible to 1 the the reverse move acceptance governed by $1/A_{ij}(\theta_i, \theta_j)$ will also be maximised resulting in easier between model moves. In general, if the move from $(M_i, \theta_i) \Rightarrow (M_j, \theta_j)$ involves a change in dimension, the best values of the parameters for the densities $q(u)$ and $q(v)$ in Equation (5) will generally be unknown, even if their structural forms are known. Using some known point $(\widetilde{u}, \widetilde{v})$, which we call the centering point, we can solve $A_{ij}(\theta_i, \theta_j) = 1$ to get the parameter values for these densities. Setting $A_{ij} = 1$ at some chosen centering point is called the zeroth-order method. Where more degrees of freedom are required we can expand A_{ij} as a Taylor series about $(\widetilde{u}, \widetilde{v})$ and solve for the proposal parameters. For the methods we use in this paper the new parameters are proposed so that the mapping function in Equation (4) is the identity function, i.e.,

$$(\theta_j, v) = h_{ij}(\theta_i, u) = (u, \theta_i)$$

and the acceptance ratio term $A_{ij}(\theta_i, \theta_j)$ probability in Equation (5) becomes

$$
\begin{aligned}
A_{ij}(\theta_i, \theta_j) &= \frac{\pi(M_j, \theta_j)}{\pi(M_i, \theta_i)} \frac{r_{ji}}{r_{ij}} \frac{q(v)}{q(u)} \\
&= \frac{\pi(M_j, \theta_j)}{\pi(M_i, \theta_i)} \frac{r_{ji}}{r_{ij}} \frac{q(\theta_i)}{q(\theta_j)}.
\end{aligned}
$$

4.2. Convergence Assessment

Convergence assessment for trans-dimensional algorithms are still in their infancy. Brooks and Giudici (1999) propose to run $I \geq 2$ chains in parallel and base their convergence diagnostic on splitting the total variation not just between chains but also between models. Their method was extended by Brooks et al. (2003) to include non-parametric techniques, including chi-square tests, Kolmogorov–Smirnov tests and direct convergence rate estimation. The latter being similar to the ideas of Raftery and Lewis (1992) for the fixed dimensional Metropolis–Hastings or Gibbs algorithms. Castelloe and Zimmerman (2002) also develop methods based on the ideas of Brooks et al. (2003) which can be used only where the parameters have the same interpretation across all models.

Brooks et al. (2003) suggest several methods for assessing convergence within the context of model selection problems. In particular for reversible jump algorithms we can have some idea of how fast the simulations approach stationarity by comparing the empirical stationary distribution on the observed model orders. They propose specific test statistics based on the χ–square distribution and also a Kolmogorov–Smirnov test for goodness of fit. The χ–square and Kolmogorov–Smirnov compare the stationary distribution of each chain and computes p–values for the computed test statistics. A critical value of 5% is used so that if the χ–square or Kolmogorov–Smirnov statistic is above this significance level there is no reason to reject the chains as not being from the same stationary distribution. See Brooks et al. (2003) for further details.

5. Model Selection Using Reversible Jump Algorithms

In this section we introduce two additional models and describe a reversible jump model selection technique to discriminate between them. Denote the full model in Equations (1) and (2) with M_1, we introduce two additional models, which are sub-models of M_1. The second model, M_2, has ρ fixed at 1. For this model there is no η and the first two levels are

$$
\begin{aligned}
R_j | \alpha'_j, \sigma &\sim \mathcal{N}\left(\alpha'_j, (\sigma' E_j)^{-1}\right) \\
\alpha'_j | \alpha'_{j-1}, \tau' &\sim \mathcal{N}\left(\alpha'_{j-1}, \tau'^{-1}\right).
\end{aligned}
$$

The prior distribution on α_0, σ and τ remain as in M_1. The posterior conditionals are exactly the same as in Section 3, simplified with $\rho = 1$ where necessary. The third model, M_3, has ρ fixed as well, however this time at 0, which results in a simple random effects model:

$$
\begin{aligned}
R_j | \alpha''_j, \sigma'' &\sim \mathcal{N}\left(\alpha''_j, (\sigma'' E_j)^{-1}\right) \\
\alpha''_j | \eta'', \tau'' &\sim \mathcal{N}\left(\eta'', \tau''^{-1}\right).
\end{aligned}
$$

The prior distributions on η, σ and τ remain as in M_1. Again the posterior conditionals are as those in Section 3 with $\rho = 0$ where necessary.

The computation here is a simple extension to the Bayesian posterior distribution described in Section 3 above. Here we have model space $\mathcal{M} = \{M_1, M_2, M_3\}$ with three models, where M_1 is the original model described in Equations (1) and (2). Models M_2 and M_3 correspond to the simplified models with ρ fixed at 1 and 0, respectively. We can extend our posterior distribution to consider both parameter and model space by taking as our posterior for model M_1

$$\pi(M_1, \alpha, \tau, \sigma, \rho, \eta | R^n) \propto L(R^n | \alpha, \sigma) p(\alpha | \rho, \eta, \tau, \alpha_0) p(\sigma) p(\tau) p(\rho) p(\eta) p(\alpha_0) p(M_1).$$

For the simplified models M_2 and M_3 the posteriors defined up to the constant of proportionality are

$$\pi(M_2, \alpha', \tau', \sigma' | R^n) \propto L(R^n | \alpha', \sigma') p(\alpha' | \tau', \alpha_0') p(\sigma') p(\tau') p(\alpha_0') p(M_2),$$

and

$$\pi(M_3, \alpha'', \tau'', \sigma'', \eta'' | R^n) \propto L(R^n | \alpha'', \sigma'') p(\alpha'' | \eta'', \tau'') p(\sigma'') p(\tau'') p(\eta'') p(M_3),$$

respectively, where $p(M_i)$ is some discrete prior distribution on the model space $\mathcal{M}$. Posterior model probabilities may then be obtained by marginalisation i.e., integrating out α, α_0, ρ, η, σ and τ to obtain the posterior marginal for M_i given the data. For the implementation we start with each model having equal prior probability

$$p(M_1) = p(M_2) = p(M_3) = \tfrac{1}{3},$$

and r_{ij} the probability of proposing a move to model M_j when at model M_i taken to be $\tfrac{1}{2}$ for $i, j = 1, 2, 3$ and $i \neq j$.

In the discussion that follows for ease of notation we suppress the dependence of the densities on the parameters $\alpha = (\alpha_1, \ldots, \alpha_n)$, σ, and τ since these parameters are common to all models. In addition for our reversible jump moves these common parameters are kept fixed between models.

5.1. Pilot Tuned Methods

Consider a proposed move from (M_2, α_0') to $(M_1, \alpha_0, \rho, \eta)$, we need to increase the dimensionality of the parameter vector by adding three components α_0, ρ and η and removing α_0'. To achieve this we simulate u_1, u_2 and u_3 from densities $q(u_1)$, $q(u_2)$ and $q(u_3)$, respectively, and set

$$(\alpha_0, \rho, \eta, v) = h_{21}(\alpha_0', u_1, u_2, u_3) = (u_1, u_2, u_3, \alpha_0'),$$

where the variable v is needed to ensure dimension matching and reversibility. We further assume v has some density $q(v)$, which we use to simulate values of v for the reverse move from M_1 to M_2. The acceptance

probability for such a move is then $\min\{1, A_{21}\}$ where

$$
\begin{aligned}
A_{21} &= \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_2, \alpha_0')} \times \frac{q(v)}{q(u_1)q(u_2)q(u_3)} \times \left| \frac{\partial h_{21}(\alpha_0', u_1, u_2, u_3)}{\partial(\alpha_0', u_1, u_2, u_3)} \right| \\
&= \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_2, \alpha_0')} \times \frac{q(v)}{q(u_1)q(u_2)q(u_3)} \\
&= \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_2, \alpha_0')} \times \frac{q(\alpha_0')}{q(\alpha_0)q(\rho)q(\eta)},
\end{aligned}
\tag{6}
$$

since the Jacobian term $\left| \frac{\partial h_{21}(\alpha_0', u_1, u_2, u_3)}{\partial(\alpha_0', u_1, u_2, u_3)} \right|$ evaluates to 1.

The densities $q(\alpha_0)$, $q(\rho)$, $q(\eta)$ and $q(\alpha_0')$ are all assumed to be Gaussian densities, with respective parameters (m_1, σ_1), (m_2, σ_2), (m_3, σ_3) and (m_v, σ_v). Theoretically, we can choose arbitrary values for the location parameters m_1, m_2, m_3 and m_v and for the scale parameters σ_1, σ_2, σ_3 and σ_v. However, some choices will result in an algorithm which takes longer to reach stationarity, since poor choices will result in low acceptance rates for between model moves. We fine-tuned the between model transitions by trying several different choices for these quantities and all resulted in the same posterior model probabilities. Generally, picking m_1 and σ_1 close to the posterior marginal mean and variance for α_0; m_2 and σ_2 close to the marginal posterior mean and variance of ρ; m_3 and σ_3 close to the marginal mean and variance of η ; m_v and σ_v close to the posterior marginal mean and variance of α_0' results in an algorithm where between model jumps are easier. We determined these posterior values by running each model in turn and recording posterior estimates of the mean and variance of the model parameters. These estimates are then used as proposal parameters in the reversible jump implementation. This scheme can only be used when there are a small number of candidate models as it becomes infeasible when the number of candidate models is large. In Section 5.2 we propose to use an automatic sampler which can choose location and scale parameters to maximise between model transitions based on methods presented in Brooks et al. (2003).

The reverse move from $(M_1, \alpha_0, \rho, \eta)$ to (M_2, α_0') is achieved by simulating v from density $q(v)$ then setting

$$
(\alpha_0', u_1, u_2, u_3) = h_{21}^{-1}(\alpha_0, \rho, \eta, v) = (v, \alpha_0, \rho, \eta)
$$

for which the acceptance probability of accepting this dimension changing move is then $\min\{1, A_{21}^{-1}\}$ where A_{21} is given in Equation (6).

The description is similar for moves between models M_1 and M_3. Assume we are at model (M_3, η'') and a move to model M_1 is proposed. We simulate u_1, u_2, u_3 from densities $q(u_1)$, $q(u_2)$ and $q(u_3)$ respectively and set

$$
(\alpha_0, \rho, \eta, w) = h_{31}(\eta'', u_1, u_2, u_3) = (u_1, u_2, u_3, \eta''),
$$

where w is introduced to ensure dimension matching.

The probability of accepting this move is then $\min\{1, A_{31}\}$ where

$$
A_{31} = \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_3, \eta'')} \times \frac{q(w)}{q(u_1)q(u_2)q(u_3)} \times \left| \frac{\partial h_{31}(\eta'', u_1, u_2, u_3)}{\partial(\eta'', u_1, u_2, u_3)} \right|.
\tag{7}
$$

For reasons similar to those given above $q(u_1)$, $q(u_2)$, $q(u_3)$ and $q(w)$ are densities approximating the poste-

Table 3. Parameter Estimates and 95% HPD Intervals. The corresponding results for the full model are given in Table 1.

	M_2	95% HPD Interval	M_3	95% HPD Interval
α_0	0.0252	(-0.0586, 0.1092)	-	
α_1	0.0253	(-0.0185, 0.0700)	0.0275	(-0.0145, 0.0697)
α_2	0.0253	(-0.0129, 0.0648)	0.0244	(-0.0170, 0.0666)
α_3	0.0368	(-0.0015, 0.0744)	0.0403	(-0.0024, 0.0818)
α_4	0.0292	(-0.0073, 0.0664)	0.0261	(-0.0143, 0.0670)
α_5	0.0358	(-0.0004, 0.0720)	0.0359	(-0.0048, 0.0754)
α_6	0.0362	(-0.0003, 0.0726)	0.0361	(-0.0032, 0.0747)
α_7	0.0304	(-0.0127, 0.0740)	0.0288	(-0.0127, 0.0706)
η	-		0.0313	(-0.0014, 0.0636)
σ	1145.7	(0.18, 2884.4)	1115.2	(1.46, 2695.4)
τ	1460.8	(35.6, 3359.3)	1617.1	(53.7, 3783.7)

rior marginals of α_0, ρ, η and η'', respectively. The reverse move from $(M_1, \alpha_0, \rho, \eta) \Rightarrow (M_3, \eta'')$ is achieved by simulating w with density $q(w)$ and setting

$$(\eta'', u_1, u_2, u_3) = h_{31}^{-1}(\alpha_0, \rho, \eta, w) = (w, \alpha_0, \rho, \eta)$$

for which the acceptance probability is the $\min\{1, A_{31}^{-1}\}$ where A_{31} is given in Equation 7.

For a proposed move from (M_2, α'_0) to (M_3, η''), we simulate w with density $q(w)$ and set $(\eta'', v) = h_{23}(\alpha'_0, w) = (w, \alpha'_0,)$. For such a proposal the acceptance probability is $\min\{1, A_{23}\}$ where

$$A_{23} = \frac{\pi(M_3, \eta'')}{\pi(M_2, \alpha'_0)} \times \frac{q(v)}{q(w)} \times \left| \frac{\partial h_{32}(\alpha'_0, w)}{\partial(\alpha'_0, w)} \right|$$

where $q(w)$ and $q(v)$ are the densities discussed above. Again the Jacobian for this proposed move is 1 since the transformation from (M_2, α'_0) to (M_3, η'') is the identity function. Notice that with this move we are not changing the number of parameters, but swapping α'_0 for η''. The acceptance probability for the reverse move is then $\min\{1, A_{23}^{-1}\}$.

5.1.1. Simulation Study

To test how well the model discrimination scheme works we simulated several datasets and applied the algorithm to them. In all cases where data were simulated from model M_2 the algorithm placed the largest posterior probability on that model, like with data simulated from model M_3 the algorithm placed the highest posterior probability on that model. For data simulated from model M_1 in some instances the highest posterior probability is placed on either model M_2 or model M_3. As the value of n increases, it appears as though the algorithm will place most of the posterior probabilities on either M_2 or M_3 since for large values of n the values of R_j simulated approach η asymptotically, hence the smaller models M_2 and M_3 offer a better fit to the data.

5.1.2. Model Averaged Results

The posterior parameter estimates with 95% HPD intervals for each of the three models are given in Table 3. The posterior model probabilities are shown in Figure 7, this shows that model M_2 has the greatest posterior
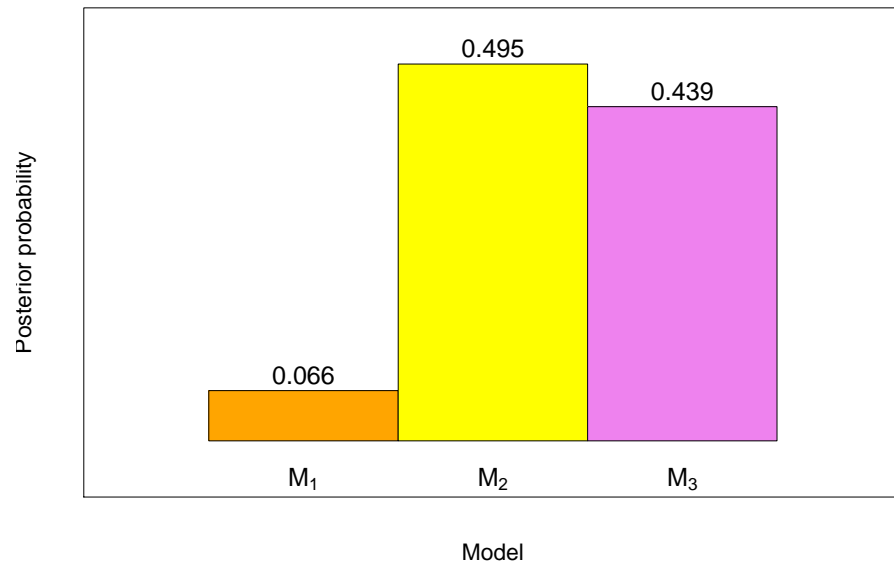
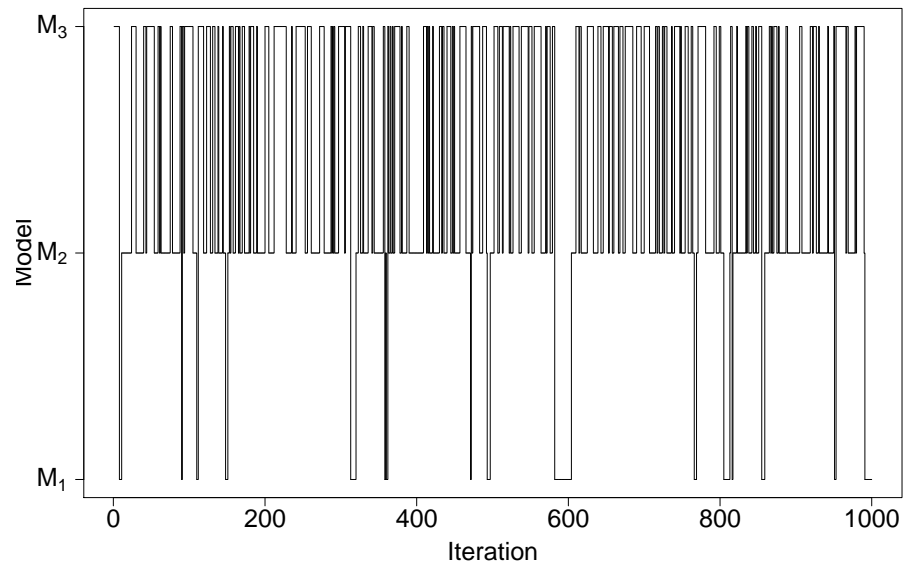**Fig. 7.** Posterior model probabilities for models M_1, M_2, and M_3.

Fig. 8. Trace plot of the model indicator.

probability of 0.495, followed by M_3 with probability 0.439. The full model M_1 has the least posterior probability, 0.066. The posterior model probabilities of M_2 and M_3 seem to contradict the results if we consider the posterior distribution of ρ. Figure 2 shows that the posterior density of ρ clearly has most mass around the node $\rho = 0$, so we might expect model M_3 to have the greater posterior probability.

It is interesting to note that many of the parameter estimates are similar under all three models. In particular the error variances seem to take very similar values under all three models. Thus model-averaged estimates look very similar from those derived from just a single model for this example. Note also the posterior distribution for ρ in the full model has a posterior mean of 0.220. This might naively be interpreted as suggesting that the ratio of model probabilities between Model M_2 and M_3 should be roughly $1 : 4$ rather than the $1 : 1$ ratio observed. The posterior density of rho is shown in Figure 2.

Figure 8 shows the mixing of the deterministic proposal reversible jump algorithm. It is noticeable that even though models M_2 and M_3 have approximately equal posterior probabilities the algorithm does not mix very well. In the next section we set try to improve the mixing of the reversible jump algorithm.

5.2. Automatic Proposal Choices

The choice of proposal densities in the reversible jump MCMC implementations given in Section 5 are determined by doing a pilot run to determine good parameter choices to describe these densities. In this section we show how this process can be made more automatic by proposing an adaptive scheme where the proposals are chosen as to maximise the probability of between model moves. Automatic proposals are desirable for a number of reasons, mainly because they reduce the need to do trial runs in order to get parameter estimates for proposal densities. The method we use is based on Brooks et al. (2003) and uses the idea of so-called weak non-identifiability and centering to determine the choice of proposal densities which maximises the probability of between model moves. The weak non-identifiability centering point is a choice of parameter values which essentially reduces the more complex model to the simpler model. We refer to this new implementation as the efficient proposals method and the previous implementation in Section 5 as the vanilla implementation. In the remainder of this section we show the details of how the between model moves are implemented.

5.2.1. Moving between Models M_1 and M_2

Consider a move from model M_2 to model M_1. The acceptance probability for such a move is $\min\{1, A_{21}\}$, where

$$A_{21} = \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_2, \alpha_0')} \frac{q(\alpha_0')}{q(\alpha_0, \rho, \eta)}. \tag{8}$$

An ideal choice for $q(\alpha_0, \rho, \eta)$ would be $\pi(\alpha_0, \rho, \eta | M_1)$, the conditional posterior for (α_0, ρ, η) given $M = M_1$. This density is non-standard, furthermore we would also need to know its normalising constant to compute the ratio A_{21}. We cannot sample directly from this density, but instead we approximate $q(\alpha_0, \rho, \eta)$ with a trivariate normal density. We approximate $q(\alpha_0')$ using a Gaussian density whose parameters we derive below. Similar methods have been proposed (Carlin and Chib, 1995; Madigan and York, 1995).

The best approximating density for $q(\alpha_0, \rho, \eta)$ in our case is one that will maximise A_{21}. To do this we use the kth-order method of Brooks et al. (2003) and expand A_{21} as a Taylor series around some point $(\tilde{\alpha}_0, \tilde{\rho}, \tilde{\eta})$ which they call the centering point. Since we need only to estimate the mean and variance of this

trivariate normal density, partial derivatives of order 1 and 2 will suffice. Essentially this means solving

$$\frac{\partial^k}{\partial(\alpha_0,\rho,\eta)^k}A_{21}\Big|_{(\tilde{\alpha}_0,\tilde{\rho},\tilde{\eta})}=0, \quad k=1,2$$

for the mean vector and covariance matrix for the density $q(\alpha_0,\rho,\eta)$, where $(\tilde{\alpha}_0,\tilde{\rho},\tilde{\eta})$ is our chosen centering point. However it is usually much easier to do computations with the log of A_{21}, in which case we solve

$$\frac{\partial^k}{\partial(\alpha_0,\rho,\eta)^k}\log A_{21}\Big|_{(\tilde{\alpha}_0,\tilde{\rho},\tilde{\eta})}=0, \quad k=1,2. \tag{9}$$

With A_{21} as given in (8) it is not very difficult to see that when we take derivatives of A_{21} (or $\log A_{21}$) with respect to (α_0,ρ,η) the terms involving $\pi(M_2,\alpha_0')$ and $q(\alpha_0')$ will contribute nothing to that derivative and likewise when we take the derivative of A_{21} or $(\log A_{21})$ with respect to α_0' the terms $\pi(M_1,\alpha_0,\rho,\eta)$ and $q(\alpha_0,\rho,\eta)$ will contribute nothing to that derivative. In what follows we will ignore terms where appropriate. Thus we can compute the first and second partial derivatives of $\log A_{21}$ as

$$\frac{\partial\log A_{21}}{\partial(\alpha_0,\rho,\eta)}=\frac{\partial}{\partial(\alpha_0,\rho,\eta)}\Big(\log\pi(M_1,\alpha_0,\rho,\eta)-\log q(\alpha_0,\rho,\eta)+K_2\Big) \tag{10}$$

and

$$\frac{\partial^2\log A_{21}}{\partial(\alpha_0,\rho,\eta)^2}=\frac{\partial^2}{\partial(\alpha_0,\rho,\eta)^2}\Big(\log\pi(M_1,\alpha_0,\rho,\eta)-\log q(\alpha_0,\rho,\eta)+K_2\Big), \tag{11}$$

where the term $K_2=-\log\pi(M_2,\alpha_0')+\log q(\alpha_0')$ is independent of (α_0,ρ,η). Also we can expand the posterior density of (M_1,α_0,ρ,η)

$$\pi(M_1,\alpha_0,\rho,\eta)\propto L(R^n|\alpha,\sigma)p(\alpha|\rho,\eta,\alpha_0,\tau)p(\eta)p(\rho)p(\alpha_0)p(\sigma)p(\tau)p(M_1)$$

and the proposal density for (α_0,ρ,η)

$$q(\alpha_0,\rho,\eta)\propto|\Sigma|^{-3/2}\exp\left\{-\tfrac{1}{2}\left(\begin{pmatrix}\alpha_0\\\rho\\\eta\end{pmatrix}-\mu\right)'\Sigma^{-1}\left(\begin{pmatrix}\alpha_0\\\rho\\\eta\end{pmatrix}-\mu\right)\right\},$$

the density of a trivariate normal distribution with mean vector μ and covariance matrix Σ. Setting (10) and (11) equal to zero at the point $(\tilde{\alpha}_0,\tilde{\rho},\tilde{\eta})$ we get two equations which can be solved simultaneously for the variance matrix Σ and mean vector μ. Solving simultaneously we can easily see that the variance matrix Σ is

$$\Sigma^{-1}=$$

$$\begin{pmatrix}1+\tau\tilde{\rho}^2 & -\tau(\alpha_1-\tilde{\eta}+2\tilde{\rho}(\tilde{\eta}-\tilde{\alpha}_0)) & -\tau\tilde{\rho}(1-\tilde{\rho})\\ -\tau(\alpha_1-\tilde{\eta}+2\tilde{\rho}(\tilde{\eta}-\tilde{\alpha}_0)) & 1+\tau\sum_{j=1}^n(\tilde{\eta}-\alpha_{j-1})^2 & -\tau\sum_{j=1}^n[(1-2\tilde{\rho})(\tilde{\eta}-\alpha_{j-1})+\tilde{\eta}-\alpha_j]\\ -\tau\tilde{\rho}(1-\tilde{\rho}) & -\tau\sum_{j=1}^n[(1-2\tilde{\rho})(\tilde{\eta}-\alpha_{j-1})+\tilde{\eta}-\alpha_j] & 1+n\tau(1-\tilde{\rho})^2\end{pmatrix} \tag{12}$$

and the mean vector μ satisfies

$$\Sigma^{-1}\left(\begin{pmatrix}\tilde{\alpha}_0\\\tilde{\rho}\\\tilde{\eta}\end{pmatrix}-\mu\right)=\begin{pmatrix}\tilde{\alpha}_0-\tilde{\rho}\tau(\alpha_1-\tilde{\rho}\tilde{\alpha}_0-(1-\tilde{\rho})\tilde{\eta})\\ \tilde{\rho}+\tau\sum_{j=1}^n[\tilde{\eta}-\alpha_{j-1}][(\tilde{\eta}-\alpha_{j-1})\tilde{\rho}+\alpha_j-\tilde{\eta}]\\ \tilde{\eta}-\tau(1-\tilde{\rho})\sum_{j=1}^n[\alpha_j-\tilde{\rho}\alpha_{j-1}-(1-\tilde{\rho})\tilde{\eta}]\end{pmatrix}$$

which results in the estimate

$$\mu = \begin{pmatrix} \tilde{\alpha}_0 \\ \tilde{\rho} \\ \tilde{\eta} \end{pmatrix} - \Sigma \begin{pmatrix} \tilde{\alpha}_0 - \tilde{\rho}\tau(\alpha_1 - \tilde{\rho}\tilde{\alpha}_0 - (1-\tilde{\rho})\tilde{\eta}) \\ \tilde{\rho} + \tau \sum_{j=1}^{n} [\tilde{\eta} - \alpha_{j-1}][(\tilde{\eta} - \alpha_{j-1})\tilde{\rho} + \alpha_j - \tilde{\eta}] \\ \tilde{\eta} - \tau(1-\tilde{\rho}) \sum_{j=1}^{n} [\alpha_j - \tilde{\rho}\alpha_{j-1} - (1-\tilde{\rho})\tilde{\eta}] \end{pmatrix}. \tag{13}$$

A difficulty arises however, since the above inverse variance matrix is not guaranteed to be positive definite (symmetric yes!) as the elements are random. Essentially, this means that the derivatives are not zero within the range of positive definite matrices, Σ. On average in this implementation Σ fails to be positive definite every 16 iterations. Our approach will be to use (12) when it is positive definite.

In cases where (12) is not positive definite we force the off-diagonal elements to be zero. Note that forcing the off-diagonal elements to being identically zero reduces our proposal from being a trivariate normal to being a product of three univariate normals. There are two possible centering points if the off-diagonal elements are set to 0, corresponding to $\tilde{\rho} = 0$ or $\tilde{\rho} = 1$. We pick the one corresponding to $\tilde{\rho} = 1$ since M_2 is a sub-model of M_1 with ρ identically equal to 1. Also with $\tilde{\rho} = 1$ fixing the off-diagonal elements at 0 dictates that $\tilde{\alpha}_0 = \alpha_7$ and $\tilde{\eta} = 2\alpha_7 - \alpha_1$.

To get the parameters for the density $q(\alpha_0')$ we simply use the conditional posterior of α_0 given $M = M_2$. This density has mean $(1+\tau)^{-1}(\tau\alpha_1)$ and variance $(1+\tau)^{-1}$. This choice can be shown to be optimal in terms of maximising the acceptance probability for proposed moves and also satisfies the kth-order equations (9). To see this, we expand

$$\pi(M_2, \alpha_0') \propto L(R^n | \alpha', \sigma') p(\alpha' | \alpha_0') p(\alpha_0') p(\sigma') p(\tau') p(M_2)$$

and supposing that $q(\alpha_0') \sim \mathcal{N}(\mu_0', v_0')$, we compute the equations

$$\frac{\partial}{\partial \alpha_0'} \log A_{21} \bigg|_{\tilde{\alpha}_0'} = \frac{\partial}{\partial \alpha_0'} \left(\log q(\alpha_0') - \log \pi(M_2, \alpha_0') + K_1 \right) \bigg|_{\tilde{\alpha}_0'} = 0,$$

$$\frac{\partial^2}{\partial (\alpha_0')^2} \log A_{21} \bigg|_{\tilde{\alpha}_0'} = \frac{\partial^2}{\partial (\alpha_0')^2} \left(\log q(\alpha_0') - \log \pi(M_2, \alpha_0') + K_1 \right) \bigg|_{\tilde{\alpha}_0'} = 0.$$

The term $K_1 = -\log \pi(M_1, \rho, \alpha_0, \eta) + \log q(\rho, \alpha_0, \eta)$ is independent of the parameter of interest α_0'. Solving simultaneously leads to the estimates $\mu_0' = (1+\tau)^{-1}(\tau\alpha_1)$ and $v_0' = (1+\tau)^{-1}$ for the mean and variance of the proposal distribution. These values are independent of the centering point $\tilde{\alpha}_0'$ chosen.

Note that when $\tilde{\rho} = 1$ the new value of η is simulated from the prior density of η, likewise when $\tilde{\rho} = 0$ α_0 is simulated from the prior density on α_0. This is a form of the birth death method for reversible jump algorithm. See Green (1995) and Brown (2004, Chapter 8).

5.2.2. Moving between Models M_1 and M_3

Consider the ratio

$$A_{31} = \frac{\pi(M_1, \alpha_0, \rho, \eta)}{\pi(M_3, \eta'')} \frac{q(\eta'')}{q(\alpha_0, \rho, \eta)},$$

notice that in taking logs and differentiating with respect to (α_0, ρ, η) we remove all terms involving M_2 and η'. For this reason the expressions given for the inverse variance matrix and mean vector for a proposed move of type M_3 to M_1 are exactly the same as those given in Equations (12) and (13). The principal difference

is that since model M_3 is a sub-model of M_1 with ρ identically equal to 0, we choose a centering point with $\tilde{\rho} = 0$. Also whenever the proposed variance matrix is not positive definite we again force the off-diagonal elements to be zero which forces $\tilde{\alpha}_0 = 2n\alpha_1 - 2\sum_{j=1}^{n}\alpha_j + \alpha_7$ and $\tilde{\eta} = \alpha_1$. Likewise the parameters for the proposal density $q(\eta'')$ which maximises A_{31} can be shown to be the posterior conditional mean of η'' and the posterior conditional variance of η'' given that $M = M_3$. This density has mean $(1+n\tau)^{-1}(\tau\sum_{j=1}^{n}\alpha_j)$ and variance $(1+n\tau)^{-1}$.

5.2.3.   Moving between Models M_2 and M_3

For a move between models M_2 and M_3 there is no change in the size of the parameter vector. The acceptance probability for such a move is $\min\{1, A_{32}\}$ where

$$A_{32} = \frac{\pi(M_2, \alpha_0')}{\pi(M_3, \eta'')} \frac{q(\eta'')}{q(\alpha_0')}.$$

We use Gaussian densities for the proposals $q(\eta'')$ and $q(\alpha_0')$. Solving

$$\left.\frac{\partial}{\partial \eta''} \log A_{32}\right|_{\tilde{\eta}''} = 0 \text{ and } \left.\frac{\partial^2}{\partial (\eta'')^2} \log A_{32}\right|_{\tilde{\eta}''} = 0$$

simultaneously shows that $q(\eta'')$ has mean $(1+n\tau)^{-1}(\tau\sum_{j=1}^{n}\alpha_j)$ and variance $(1+n\tau)^{-1}$. The reader will notice at once that these quantities are the conditional posterior mean and variance of η'' given $M = M_3$. Similarly solving

$$\left.\frac{\partial}{\partial \alpha_0'} \log A_{32}\right|_{\tilde{\alpha}_0'} = 0 \text{ and } \left.\frac{\partial^2}{\partial (\alpha_0')^2} \log A_{32}\right|_{\tilde{\alpha}_0'} = 0$$

simultaneously shows that $q(\alpha_0')$ has mean $(1+\tau)^{-1}(\tau\alpha_1)$ and variance $(1+\tau)^{-1}$, which are the conditional posterior mean and variance of α_0' given $M = M_2$.

We can summarise this by saying that $q(\eta'') = \pi(\eta''|M_3)$ and $q(\alpha_0') = \pi(\alpha_0'|M_2)$ are the proposals which will maximise the acceptance probability for proposed moves between models M_2 and M_3, and that these choices are independent of the centering point chosen. In this case the ratio A_{32} reduces to

$$\begin{aligned} A_{32} &= \frac{\pi(M_2, \alpha_0')}{\pi(M_3, \eta'')} \frac{q(\eta'')}{q(\alpha_0')} \\ &= \frac{\pi(M_2, \alpha_0')}{\pi(M_3, \eta'')} \frac{\pi(\eta''|M_3)}{\pi(\alpha_0'|M_2)}. \end{aligned}$$

In our simulations using this term should increase the between model moves. This was observed in our simulations as all proposed moved from model M_3 to model M_2 were accepted, whereas for the vanilla implementation such moves were accepted with probability 0.498. Similarly a proposed move from model M_2 to model M_3 is accepted with probability 0.930 when the posterior conditionals are used as proposals, improving upon the 0.440 probability obtained with the vanilla implementation. The empirical results observed here are actually specific cases of more general results which can be found in Ehlers and Brooks (2002).
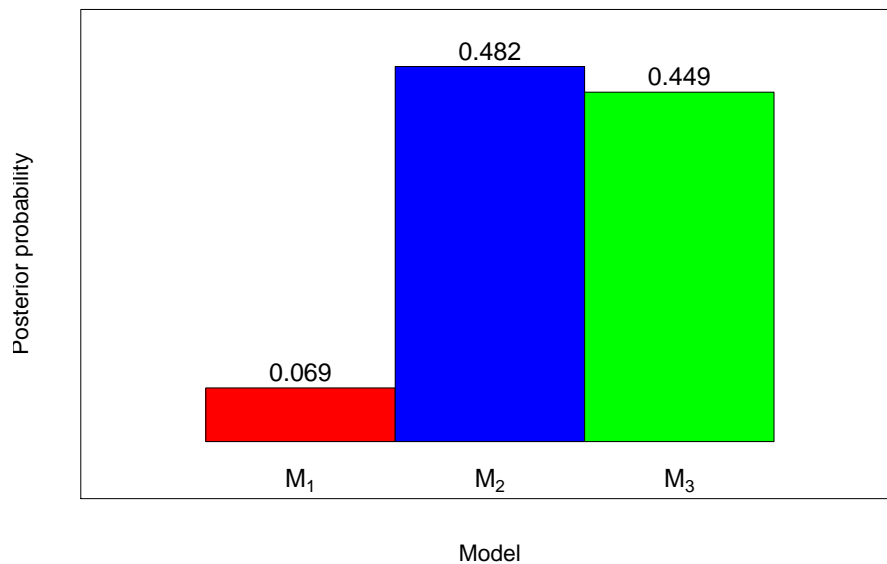
Fig. 9. Posterior model probabilities for models M_1, M_2, and M_3, second reversible jump implementation.
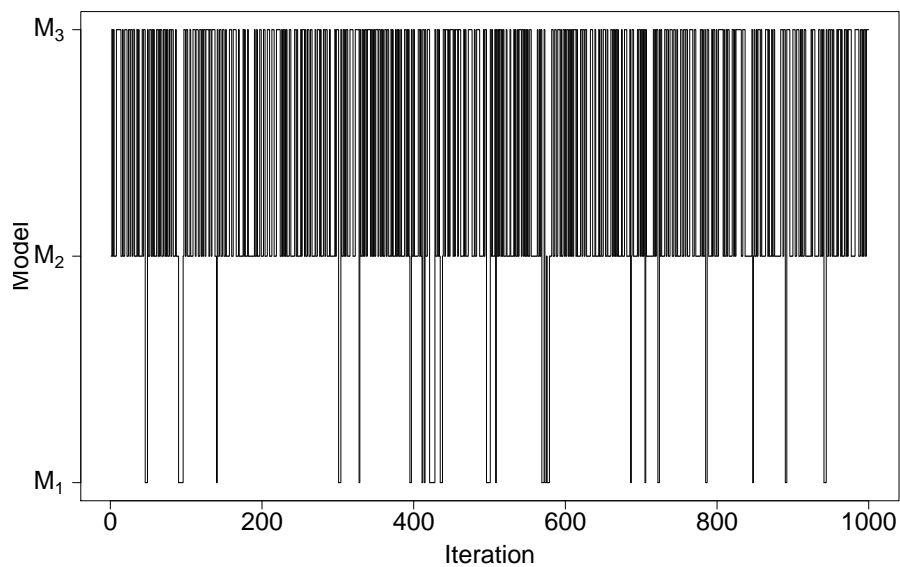

Fig. 10. Trace plot of the model indicator, second reversible jump implementation. The horizontal axis shows the iteration number and the vertical axis shows the model indicator.
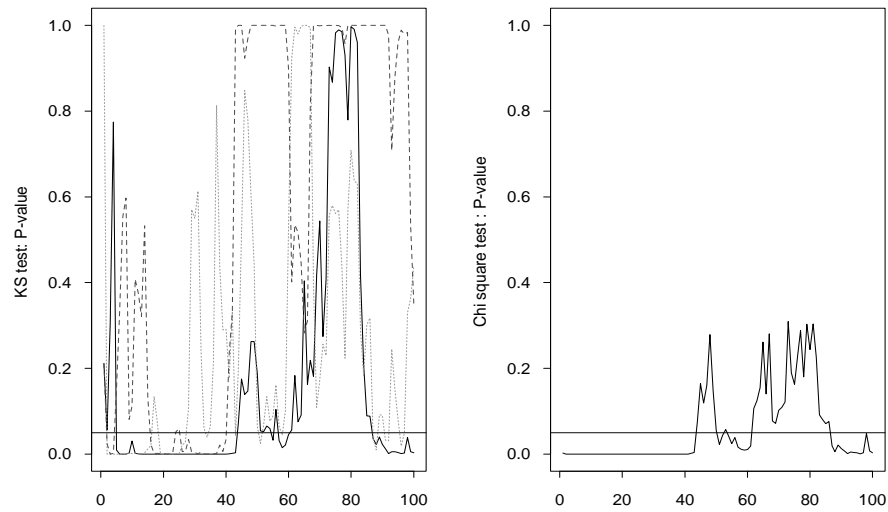
Fig. 11. Convergence diagnostics for the vanilla implementation. The horizontal axis times 1000 gives the iteration number.
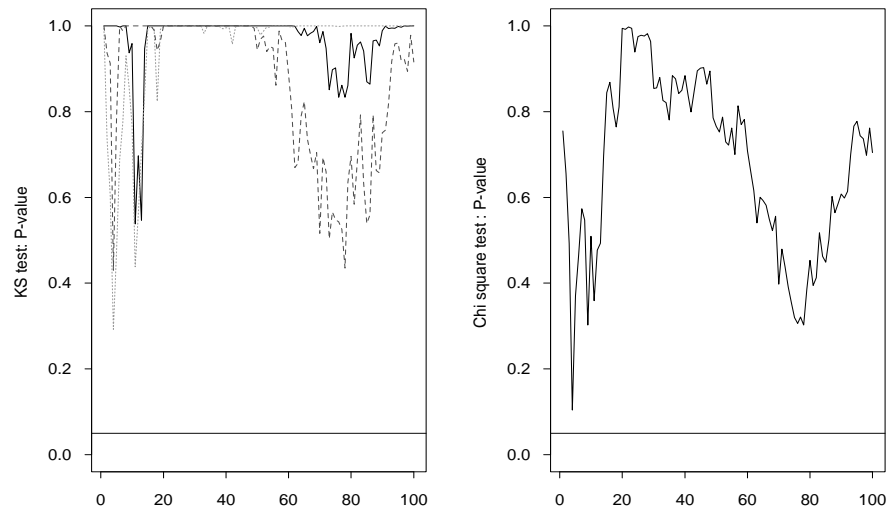

Fig. 12. Convergence diagnostics for the automatic proposals implementation. The horizontal axis times 1000 gives the iteration number.

5.3. Comparing the Model-move Schemes

The empirical transition matrices for the vanilla reversible jump method, P^{van}, and for the second order method, P^{eff}, are, respectively

$$
P^{\text{van}} = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{ccc} M_1 & M_2 & M_3 \\ \begin{pmatrix} 0.703 & 0.154 & 0.142 \\ 0.020 & 0.758 & 0.220 \\ 0.021 & 0.249 & 0.729 \end{pmatrix} \end{array} \text{ and } P^{\text{eff}} = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{ccc} M_1 & M_2 & M_3 \\ \begin{pmatrix} 0.597 & 0.121 & 0.281 \\ 0.018 & 0.516 & 0.465 \\ 0.043 & 0.501 & 0.456 \end{pmatrix} \end{array}
$$

The empirical transition matrices are computed by setting the (i, j)-element equal to the proportion of times the model indicator M_j follows the model indicator M_i for one long run of the reversible jump algorithm, in this case for 1000000 iterations.

They matrices clearly that between model (off-diagonal) transitions have increased for P^{eff} the transition matrix for the efficient proposals method, except between models M_1 and M_2 where there were small decreases. To assess convergence of the algorithm, we simulated 3 chains using different starting values and different random number seeds for a total of 1000000 iterations. In Section 4.2 we introduced two methods of assessing convergence of reversible jump chains. Both the χ-square and Kolmogorov–Smirnov diagnostics are used to assess convergence of our simulations. These diagnostics are plotted in Figures 11 and 12 for the vanilla reversible jump algorithm and efficient proposals implementations, respectively. Clearly the efficient proposals implementation performs better than the vanilla implementation

We summarise by giving the efficient proposals results applied to the models discussed in Section 5 and compare them with those obtained using the vanilla reversible jump algorithm using the fine-tuned proposals described in Section 5. We end this section by briefly addressing convergence issues. The posterior model probabilities are shown in Figure 9, the posterior model probabilities are similar to those obtained in Section 5. Model M_1 has posterior probability 0.069, M_2 has posterior probability 0.482 and M_3 has posterior probability 0.449. While the computing effort required to implement this model is a bit greater than that required for the vanilla reversible jump method, the improved mixing can also be seen by comparing Figures 8 and 10. Figure 10 shows that the algorithm jumps between models more frequently for the second implementation compared with the fine-tuned proposals implementation shown in Figure 8. The within model parameter estimates are almost identical to those obtained using the implementation in Section 5 and are not tabulated here. The minor differences we attribute to Monte Carlo errors.

6. Summary

The reversible jump algorithm is presented as a method of computing posterior model probabilities in a Bayesian setting. The vanilla reversible jump algorithm although theoretically sound has some implementational problems. One such problem is the choice of mapping function, another is the choice of proposal density parameters. In this paper we have shown how recent methodological advances in reversible jump computing can be applied to model selection problems. This is particularly useful for actuarial practitioners where the most appropriate choice of model is important.

References

Al-Awadhi, F., C. Jennison, and M. Hurn (2004). Statistical image analysis for a confocal microscopy two–dimensional section of a cartilage growth. *Journal of the Royal Statistical Society, Series C* **53**, 31–49.

Brooks, S. P. and P. Giudici (1999). Diagnosing Convergence of Reversible Jump MCMC Algorithms. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*, pp. 733–742. Oxford University Press.

Brooks, S. P., P. Giudici, and A. Philippe (2003). Nonparametric Convergence Assessment for MCMC Model Selection. *Journal of Computational and Graphical Statistics* **12**, 1–22.

Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society, Series B* **65**(1), 3–55.

Brown, G. O. (2004). *Model Discrimination in Bayesian Credibility Modelling*. Ph. D. thesis, Cambridge University.

Bühlmann, P. and H. Bühlmann (1999). Selection of credibility regression models. *ASTIN Bulletin* **29**(2), 245–270.

Cairns, A. J. G. (2000). A discussion of parameter and model uncertainly in insurance. *Insurance: Mathematics and Economics* **27**, 313–330.

Carlin, B. P. and S. Chib (1995). Bayesian Model Choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **57**, 473–484.

Castelloe, J. M. and D. L. Zimmerman (2002). Convergence Assessment for Reversible Jump MCMC Samplers. Technical report, SAS Institute, Dept. of Statistics and Actuarial Science, U. of Iowa.

Ehlers, R. S. and S. P. Brooks (2002). Efficient Construction of Reversible Jump MCMC Proposals for ARMA Models. Technical report, Universidade Federal do Parana, Department de Estatistica.

George, E. I. and R. E. McCulloch (1993). Stochastic Search Variable Selection. *Journal of the American Statistical Society* **88**, 881–889.

Giudici, P. and G. O. Roberts (1998). On the automatic choice of reversible jumps. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*. Oxford University Press.

Godsill, S. J. (2001). On the relationship between Markov chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics* **10**(2), 230–248.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4), 711–732.

Green, P. J. (2002). Trans-dimensional Markov chain Monte Carlo. In *Highly Structured Stochastic Systems*, pp. 179–198. Oxford University Press.

Green, P. J. and A. Mira (2001). Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* **88**(4), 1035–1053.

Grenander, U. and M. I. Miller (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society, Series B* **56**, 549–603.

Harris, G. (1999). Markov chain Monte Carlo estimation of regime switching vector autoregressions. *ASTIN Bulletin* **29**(1), 47–80.

Keatinge, C. L. (1999). Modeling Losses with the Mixed Exponential Distribution. *Proceedings of the Casual Actuarial Society* **LXXXV1**, 654–698.

Klugman, S. A. (1992). *Bayesian Statistics in Actuarial Science*. Boston, MA: Kluwer Academic Publishers.

Ledolter, J., S. Klugman, and C.-S. Lee (1991). Credibility models with time-varying trend components. *ASTIN Bulletin* **21**(1), 73–91.

Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63**, 215–232.

Papaspiliopoulos, O., G. O. Roberts, and M. Sköld (2003). Non-centered parameterizations for hierarchical models and data augmentation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, and A. P. Dawid (Eds.), *Bayesian Statistics*, Volume **7**, pp. 307–326. Oxford University Press.

Phillips, D. B. and A. F. M. Smith (1996). Bayesian model comparison via jump diffusions. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 215–239. Chapman and Hall.

Raftery, A. E. and S. M. Lewis (1992). How Many Iterations in the Gibbs Sampler. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger (Eds.), *Bayesian Statistics 4*, pp. 763–774. Oxford University Press.

Roberts, G. O. and J. S. Rosenthal (1998). Markov Chain Monte Carlo: Some practical implications of theoretical results. *Canadian Journal of Statistics* **26**(1), 5–20.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**(4), 351–367.

Rotondi, R. (2002). On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change–points. *Computational Statistics and Data Analysis* **40**(3), 633–653.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of Statistics* **28**(1), 40–74.

Troughton, P. T. and S. J. Godsill (1997). A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical report, Department of Engineering, University of Cambridge, Signal Processing and Communications Laboratory.